Dispersion Models for Extremes

Bent Jørgensen[1] · Yuri Goegebeur[1] · José Raúl Martínez[2]

February 2, 2008

Abstract We propose extreme value analogues of natural exponential families and exponential dispersion models, and introduce the slope function as an analogue of the variance function. The set of quadratic and power slope functions characterize well-known families such as the Rayleigh, Gumbel, power, Pareto, logistic, negative exponential, Weibull and Fréchet. We show a convergence theorem for slope functions, by which we may express the classical extreme value convergence results in terms of asymptotics for extreme dispersion models. The main idea is to explore the parallels between location families and natural exponential families, and between the convolution and minimum operations.

Keywords Convergence of extremes · Extreme dispersion model · Generalized extreme value distribution · Hazard location family · Power slope function · Quadratic slope function

AMS 2000 Subject Classification Primary—62E10, 60F99, 60E10; Secondary—62E20, 62N99

1 Introduction

In a seminal paper Morris (1982) asked the following question: what do the normal, Poisson, gamma, binomial, and negative binomial distributions have in common that makes them so special? His answer was that they are all natural exponential families with quadratic variance functions. This idea has wide-ranging practical and theoretical ramifications, in particular for generalized linear models (McCullagh and Nelder, 1989) and exponential dispersion models (Jørgensen, 1987).

Many subsequent authors have used the variance function as a characterization and convergence tool for natural exponential families and exponential dispersion models, cf. Jørgensen (1997), Casalis (2000) and references therein. In particular, Tweedie (1984) and several authors independently of him (Morris, 1981; Hougaard, 1986; Bar-Lev and Enis, 1986), proposed and investigated the class of power variance functions, corresponding to what we now call the Tweedie class of exponential dispersion models. Jørgensen et al. (1994) showed that the Tweedie models appear as limits in a class of convergence results for exponential dispersion models, extending certain classical stable convergence results.

These ideas appear, at first sight, to have little relevance for extreme value theory. Echoing Morris (1982) we may ask, however, what distributions like the Rayleigh, Gumbel, power, Pareto, logistic and negative exponential have in common that makes them so special in the context of extremes? Also, is there an extreme value analogue of power variance functions, perhaps related to the Weibull and

[1]Bent Jørgensen (corresponding author), Yuri Goegebeur
Department of Statistics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark
e-mail: bentj@stat.sdu.dk, yuri.goegebeur@stat.sdu.dk
[2]José Raúl Martínez
FAMAF, Universidad Nacional de Córdoba, Ciudad Universitaria, 5000 Córdoba, Argentina
e-mail: jmartine@mate.uncor.edu

Fréchet distributions? In the present paper we develop an extreme value dispersion model framework in the spirit of Jørgensen (1997), leading to constructive answers to these questions. In particular we find a close parallel between the above-mentioned Tweedie convergence results and the classical extreme convergence results by Fisher and Tippett (1928) and Gnedenko (1943). See Coles (2001), Kotz and Nadarajah (2002) and Beirlant et al. (2004) for background material on extremes.

In Section 2 we introduce the rate and slope of a distribution as analogues of the mean and variance, respectively. In Section 3 hazard location families and slope functions are introduced as analogues of natural exponential families and variance functions, respectively. In Section 4 we introduce extreme dispersion models as analogues of exponential dispersion models. In Section 5 we classify quadratic slope functions in a manner similar to Morris' (1982) classification of quadratic variance functions. In Section 6 we draw a parallel between generalized extreme value distributions and Tweedie models, having power slope functions and power variance functions, respectively. In Section 7 a general convergence result for slope functions is shown, leading to a new proof of the classical extreme value convergence results, now set in the extreme dispersion model setting. Finally, in Section 8 we consider characterization and convergence for exponential slope functions.

2 Basic framework

We now introduce the basic setup for the paper, define the notions of rate and slope for a real random variable Y, and show that they are analogues of the mean and variance, respectively. It is convenient to use familiar terms from lifetime analysis such as survival function and hazard function, although Y is not restricted to be positive. It is also convenient to use minimum (min) rather than the conventional maximum. Results for maxima may be obtained by a reflection in the usual way, see Beirlant et al. (2004, p. 46).

2.1 Survival, hazard and integrated hazard

We assume that the survival function $G(y) = P(Y \geq y)$ is twice continuously differentiable on the support $\mathcal{C} = (a, b) \subseteq \mathbb{R}$, continuous at a, and possibly discontinuous at b. Let $\mathcal{G}$ denote the set of all G such that the density function $f = -G'$ is strictly positive on $\mathcal{C}$, and let $\mathcal{G}_0$ denote the subset of $\mathcal{G}$ for which $0 \in \mathcal{C}$.

When $G(b) > 0$ we talk about *right censoring* at b. In particular we allow a positive probability mass $G(\infty)$ at $b = \infty$, in which case G represents an *improper* distribution. In survival analysis $G(\infty)$ is the probability that an individual never experiences the event in question, see e.g. Aalen (1988).

We define the *integrated hazard function* $H : \mathbb{R} \to [0, \infty]$ and the *hazard function* $h : \mathbb{R} \to [0, \infty]$ corresponding to G by $H(y) = -\log G(y)$ and $h(y) = H'(y)$, respectively. It is understood that both H and h are 0 to the left and ∞ to the right of $\mathcal{C}$, except that $H(b)$ is finite if $G(b) > 0$. With these conventions the following relationship holds for all $y \in \mathbb{R}$,

$$H(y) = \int_{-\infty}^{y} h(x)\, dx. \tag{1}$$

2.2 Rate, slope and semiinvariants

We now define the rate and slope for Y, and make the connection with the min operation.

By way of motivation, let us recall the derivation of the mean and variance from the moment generating function. Let the random variable Y have moment generating function $M(t) = \mathrm{E}\left(e^{tY}\right)$, with domain $\Theta = \{t \in \mathbb{R} : M(t) < \infty\}$ such that $0 \in \operatorname{int} \Theta$. Consider the *cumulant generating function* $\kappa = \log M$ whose derivatives at zero $\kappa^{(i)}(0)$ are the cumulants. The mean and variance, in particular, are given by

$$\mathrm{E}(Y) = \tau(0) \qquad \text{and} \qquad \mathrm{Var}(Y) = \tau'(0), \tag{2}$$

where $\tau = \kappa'$ is the *mean value mapping,* which is strictly increasing on intΘ.

We now propose $G(y) = \mathrm{E}\left(1_{Y \geq y}\right)$ as an analogue of the moment generating function $M(t) = \mathrm{E}\left(e^{tY}\right)$, which in turn makes H and h analogues of the cumulant generating function κ and mean value mapping τ, respectively. By analogy with (2), we define the *rate* r and *slope* s for a random variable Y with survival function $G \in \mathcal{G}_0$ by

$$r(Y) = h(0) \text{ and } s(Y) = h'(0),$$

respectively. Unlike the variance, however, the slope may be negative as well as positive, and the rate decreases (increases) under translation in the IFR (DFR) case. In general we define the ith *semiinvariant* by $k_i(Y) = H^{(i)}(0)$ for $i \geq 1$, provided the derivatives exist, analogously to the cumulants. This terminology alludes to T.N. Thiele's name *half-invariants* for the cumulants, cf. Lauritzen (2002, p. 207).

Letting $\mu = r(Y)$, the slope of Y may be written as follows:

$$s(Y) = \mu \left\{\mu - g'(0)\right\}, \tag{3}$$

where $g = -\log f$. This result is somewhat analogous to the result $\mathrm{Var}(Y) = \mathrm{E}(Y^2) - \mathrm{E}^2(Y)$ for the variance, see also Section 2.3.

Like the cumulants, the semiinvariants satisfy a scale equivariance property $k_i(cY) = c^{-i} k_i(Y)$ for $c > 0$, which follows from the fact that cY has integrated hazard function $H(y/c)$. In particular the rate and slope satisfy

$$r(cY) = c^{-1} r(Y) \text{ and } s(cY) = c^{-2} s(Y). \tag{4}$$

The rate is *not*, however, translation equivariant, nor is the slope translation invariant, but instead satisfy $r(c+Y) = h(-c)$ and $s(c+Y) = h'(-c)$ for $c \in \mathbb{R}$.

The min of n independent variables Y_i has integrated hazard function $H_1(y) + \cdots + H_n(y)$, so the semiinvariants are additive with respect to the min operation $\wedge$, in much the same way that the cumulants are additive with respect to convolution. In particular

$$r\left(\bigwedge_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} r(Y_i) \text{ and } s\left(\bigwedge_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} s(Y_i). \tag{5}$$

We denote the scaled min of n independent and identically distributed (i.i.d.) variables Y_i by

$$\hat{Y}_n = n \bigwedge_{i=1}^{n} Y_i, \tag{6}$$

which in many ways behaves like the sample mean. Thus, combining (4) and (5) yields

$$r\left(\hat{Y}_n\right) = \mu \text{ and } s\left(\hat{Y}_n\right) = \frac{\varsigma}{n}, \tag{7}$$

where μ is the rate and ς the slope of Y_i. Also, the exponential distribution is invariant under the transformation (6), behaving like a constant does under averaging. This suggest a law of large numbers involving the exponential distribution, as we shall now see.

2.3 Exponential distribution

Let E_μ denote an exponential variable with rate $\mu > 0$. By a shifted exponential variable we mean $a + E_\mu$ with $a < 0$, whose support includes 0. For such a variable we find

$$r\left(a + E_\mu\right) = \mu \text{ and } s\left(a + E_\mu\right) = 0, \tag{8}$$

parallel to the form for the mean and variance of a constant. In the notation of (3), note that $g'(0) = \mu$ for the variable $a + E_\mu$, so (3) implies that the slope is a signed measure of the deviation of Y from

exponentiality, in much the same way that the variance is a measure of the deviation of Y from being constant.

Since the exponential distribution hence plays the role of constant in the present setup, it is not surprising that there is a law of large numbers, as suggested by (7) and (8), that involves convergence to the exponential distribution. In fact, the scaled min $\hat{Y}_n$, after left truncation at 0, has survival function given, for $y > 0$, by

$$\frac{G^n\left(y/n\right)}{G^n\left(0\right)} = \exp\left\{-y\mu - \frac{\varsigma}{2n}y^2 + o(n^{-1})\right\} \text{ as } n \to \infty, \tag{9}$$

converging to an exponential distribution with rate μ. Here left truncation at 0 means conditioning on the event $\hat{Y}_n > 0$. The quadratic term suggests a central limit theorem. By removing the term $-y\mu$, corresponding to an exponential component, and rescaling we obtain for $y > 0$

$$\frac{G^n\left(y/\sqrt{n}\right)}{G^n\left(0\right)}e^{y\mu\sqrt{n}} = \exp\left\{-\frac{\varsigma}{2}y^2 + o(1)\right\} \text{ as } n \to \infty. \tag{10}$$

Provided $\varsigma > 0$, this gives an asymptotic Rayleigh distribution, which hence plays the role of the normal distribution in the present setup. Remark 5.1 below makes precise the idea of removing an exponential component.

3 Hazard location families

3.1 Motivation

We now introduce hazard location families, and show that each such family is characterized by its slope function, just like a natural exponential family is characterized by its variance function.

Recall that the *variance function* is defined on $\Omega = \tau(\operatorname{int}\Theta)$ by $V(\mu) = \tau'\left(\tau^{-1}(\mu)\right)$, where τ is the mean value mapping defined in connection with (2). As pointed out by Morris (1982), V characterizes is the distribution of Y up to an exponential tilting. This follows since given V, the inverse τ^{-1} satisfies the differential equation

$$\frac{d\tau^{-1}(\mu)}{d\mu} = \frac{1}{V(\mu)}, \tag{11}$$

from which $\tau^{-1}(\mu)$ may be recovered up to an additive constant $-\theta$, say, corresponding to an *exponential tilting* of the density f (not necessarily with respect to Lebesgue measure)

$$f(y;\theta) = f(y)\exp\left\{y\theta - \kappa(\theta)\right\}. \tag{12}$$

This is a natural exponential family (NEF), and (12) has mean $\mu = \tau(\theta)$ and variance $V(\mu)$, cf. Jørgensen (1997, Ch. 2).

3.2 Definition

Our analogy implies that the hazard function h should be analogous to the mean value mapping τ. In order to make the analogy complete, however, we need h, like τ, to be monotone. We thus consider from now on survival functions in $\mathcal{G}$ with *monotone hazard rate*, in the sense that h is strictly monotone on $\mathcal{C}$, either increasing (IFR) or decreasing (DFR). This subset of $\mathcal{G}$ is denoted $\overline{\mathcal{G}}$, and we let $\overline{\mathcal{G}}_0 = \overline{\mathcal{G}} \cap \mathcal{G}_0$.

Remark 3.1 *In the DFR case it is necessary that $a > -\infty$ in order for the integral (1) to converge at a. In the IFR case G is always proper, since if h is increasing on (a, ∞) then the integral (1) diverges at ∞.*

Table 1: The main quadratic hazard slope families and associated NEFs. GHS is the generalized hyperbolic secant family.

HL(μ)	$G(y)$	$\mathcal{C}$	$v(\mu)$	Ψ	NEF
Rayleigh	$\exp\left(-y^2/2\right)$	$\mathbb{R}_+$	1	$\mathbb{R}_+$	Normal
Gumbel	$\exp\left(-e^y\right)$	$\mathbb{R}$	μ	$\mathbb{R}_+$	Poisson
Uniform	$1-y$	$(0,1)$	μ^2	$(1,\infty)$	Gamma
Pareto	y^{-1}	$(1,\infty)$	$-\mu^2$	$(0,1)$	—
Logistic	$(1+e^y)^{-1}$	$\mathbb{R}$	$\mu(1-\mu)$	$(0,1)$	Binomial
Neg. exponential	$1-e^y$	$\mathbb{R}_-$	$\mu(1+\mu)$	$\mathbb{R}_+$	Neg. binomial
Cosine	$\cos y$	$(0,\pi/2)$	$1+\mu^2$	$\mathbb{R}_+$	GHS

Remark 3.2 *Consider the case $Y=\log T$, where T is a positive survival time, say. Since then $\mathcal{C}=\mathbb{R}$, making $a=-\infty$, only IFR is possible. The derivative of the hazard function for T is*

$$\frac{h'(\log t)-h(\log t)}{t^2}.$$

Hence T need not have monotone hazard rate, even though Y does. In this sense, the assumption of monotone hazard rate is less of a restriction when modelling log survival times.

We now define an analogue of the variance function. For given $G\in\overline{\mathcal{G}}$ we let $\Psi=h(\mathcal{C})$ (an open interval), and define the *slope function* $v:\Psi\to\mathbb{R}_\pm$ by

$$v(\mu)=h'\left(h^{-1}(\mu)\right). \tag{13}$$

Here v maps into $\mathbb{R}_+$ in the IFR case and into $\mathbb{R}_-$ in the DFR case. Analogously to (11) we find that the inverse hazard function h^{-1} satisfies

$$\frac{dh^{-1}(\mu)}{d\mu}=\frac{1}{v(\mu)}. \tag{14}$$

Proposition 3.1 *The slope function v with domain Ψ characterizes the location family $G(\cdot-\theta)$ with $\theta\in\mathbb{R}$ among all location families within $\overline{\mathcal{G}}$.*

Proof: Given v, the solution to (14) is $\theta+h^{-1}(\mu)$, where $\theta\in\mathbb{R}$ is arbitrary. By inversion, we obtain the hazard function $h(\cdot-\theta)$ corresponding to the location family $G(\cdot-\theta)$. ∎

To complete the analogy with natural exponential families we restrict the domain of θ to $-\mathcal{C}$, such that $G(\cdot-\theta)\in\overline{\mathcal{G}}_0$. Note that the rate and slope for $G(\cdot-\theta)$ are $\mu=h\left(-\theta\right)$ and $h'\left(-\theta\right)=v(\mu)$. This leads to the following definition.

Definition 3.1 *The* hazard location family $\{\mathrm{HL}(\mu):\mu\in\Psi\}\subseteq\overline{\mathcal{G}}_0$ *generated from $G\in\overline{\mathcal{G}}$ is defined by the family of survival functions with support $\mathcal{C}-h^{-1}(\mu)$ given by*

$$y\mapsto G\left\{y+h^{-1}(\mu)\right\}. \tag{15}$$

Note that the definition of v in (13) is independent of the representation (15) used for the family, so the slope function represents an intrinsic property of the family.

Table 1 shows some examples of hazard location families corresponding to familiar distributions, all with quadratic slope functions (polynomials of degree at most two), to be studied in Section 5. Except for the Pareto distribution, all the families in the table are IFR. These six IFR families have the same functional form for v as the variance functions for Morris' (1982) six natural exponential families. In particular, the Rayleigh distribution has constant slope function, like the variance function of the normal distribution.

3.3 Truncation and censoring

We now study the effect on v of transformations like truncation and censoring.

Left truncation at some point $c \in \mathcal{C}$ gives rise to a new hazard location model with v restricted to a subset of Ψ. Similarly, the operation of right censoring at some point $c \in \mathcal{C}$ corresponds to replacing Y by $Y \wedge c$, also known as the limited loss variable in loss modelling (Klugman et al., 2004, p. 30). This operation reduces $\mathcal{C}$ to the subset (a, c) and introduces the probability $G(c)$ in c. We summarize these considerations in a lemma.

Lemma 3.2 *Left truncation at $c \in \mathcal{C}$ corresponds to restricting the domain of v to the interval between $h(c)$ and $h(b)$. Right censoring at $c \in \mathcal{C}$ corresponds to restricting the domain of v to the interval between $h(a)$ and $h(c)$.*

In the DFR case, left truncation thus results in the domain $\Psi = (h(b), h(c))$, whereas right censoring gives the domain $(h(c), h(a))$. The lemma shows that the restriction of v to a subinterval of its domain is again the slope function for a hazard location family. When looking for a model corresponding to a given functional form for v, we may hence concentrate on the largest possible domain consistent with a survival function in $\overline{\mathcal{G}}_0$. Note, however, that restricting the domain of v to a subset of Ψ implies a change in the distributional form, because the support is changed. By comparison, restricting μ to a subinterval of Ω in a natural exponential family selects a subset of the family of distributions, without changing the distributions as such.

4 Extreme dispersion models

4.1 Motivation

We now introduce extreme dispersion models as a parallel to exponential dispersion models, and show that they satisfy a reproductive property.

Given a natural exponential family (12) with variance function $V(\mu)$, the corresponding *exponential dispersion model* $\mathrm{ED}(\mu, \lambda)$ consists of natural exponential families with variance function $\lambda^{-1}V(\mu)$ proportional to $V(\mu)$. The latter is then called the *unit variance function*. The model $\mathrm{ED}(\mu, \lambda)$ has density function of the form

$$f(y; \theta, \lambda) = f_\lambda(y) \exp\left[\lambda \left\{y\theta - \kappa(\theta)\right\}\right],$$

for a suitable function f_λ. Here $\mu = \tau(\theta)$ is the mean, in the notation of (2), and $\sigma^2 = 1/\lambda$ is the *dispersion parameter*. The index parameter λ has domain $\Lambda \subseteq \mathbb{R}_+$, which is an additive semigroup (often $\mathbb{R}_+$ or $\mathbb{N}$).

$\mathrm{ED}(\mu, \lambda)$ satisfies the following *mean reproductive property*. The average of n i.i.d. variables $Y_1, \ldots, Y_n$ from $\mathrm{ED}(\mu, \lambda)$ has distribution

$$\bar{Y}_n \sim \mathrm{ED}(\mu, n\lambda), \tag{16}$$

where the index parameter is proportional to the sample size. This follows from the form of the moment generating function of $\mathrm{ED}(\mu, \lambda)$, which is

$$t \mapsto \frac{M^\lambda(t/\lambda + \tau^{-1}(\mu))}{M^\lambda(\tau^{-1}(\mu))}, \tag{17}$$

where $M(\cdot)$ is the moment generating function for $f = f_1$, cf. Jørgensen (1997, Ch. 3).

4.2 Definition

Definition 4.1 *Given a survival function $G \in \overline{\mathcal{G}}$ with hazard function h and support $\mathcal{C}$, we define the* extreme dispersion model generated by G, *denoted* $\mathrm{XD}(\mu, \lambda)$, *as the family of survival functions in $\overline{\mathcal{G}}_0$ given by*

$$y \mapsto G^{\lambda}(y/\lambda + h^{-1}(\mu)) \tag{18}$$

with rate $\mu \in \Psi$, index parameter $\lambda > 0$ and support $\lambda\left(\mathcal{C} - h^{-1}(\mu)\right)$.

It is straightforward to see that the model $\mathrm{XD}(\mu, \lambda)$ may be generated from any of its members in this way, up to a rescaling of λ to $c\lambda$ for some $c > 0$. In the following we work with the representation (18) corresponding to a specific choice for G. The corresponding hazard and density functions are for $y \in \lambda\left(\mathcal{C} - h^{-1}(\mu)\right)$ given by

$$h(y; \mu, \lambda) = h(y/\lambda + h^{-1}(\mu)) \tag{19}$$

and

$$f(y; \mu, \lambda) = h(y/\lambda + h^{-1}(\mu)) \exp\left[-\lambda H\left\{y/\lambda + h^{-1}(\mu)\right\}\right],$$

respectively. The right extreme of the support is $\lambda(b - h^{-1}(\mu))$, which has probability $G^{\lambda}(b)$.

The parameter μ is the rate for $\mathrm{XD}(\mu, \lambda)$ for any value of $\lambda > 0$, which follows from (19) by inserting $y = 0$, in much the same way that μ is the mean of $\mathrm{ED}(\mu, \lambda)$ for all λ. For each fixed value of λ, (18) corresponds to a hazard location model with slope function $\lambda^{-1}v(\mu)$. Hence v is called the *unit slope function* for $\mathrm{XD}(\mu, \lambda)$, and by Proposition 3.1 v characterizes $\mathrm{XD}(\mu, \lambda)$ up to a rescaling of λ like above. We call $\sigma^2 = 1/\lambda$ the *dispersion parameter*.

The following *min reproductive property* easily follows from the form of the survival function (18). For i.i.d. variables $Y_1, \ldots, Y_n \sim \mathrm{XD}(\mu, \lambda)$ the scaled min $\hat{Y}_n$ from (6) has distribution

$$\hat{Y}_n \sim \mathrm{XD}(\mu, n\lambda). \tag{20}$$

This is analogous to the mean reproductive property (16) for exponential dispersion models. It is equivalent to the max-stable property of an exponentiated family of distributions (Nelson and Doganaksoy, 1995; Sarabia and Castillo, 2005; Nadarajah and Kotz, 2006), but the present formulation emphasizes the fact that the rate is preserved under the scaled min operation. In (20), like (16), the index parameter is proportional to the sample size.

Let us consider the $\mathrm{XD}(\mu, \lambda)$ models generated from first six cases in Table 1, where in fact the introduction of the index parameter λ corresponds to known generalizations. In the Rayleigh and Gumbel cases, this adds a scale or location parameter to the models, respectively. The uniform distribution becomes a shifted power distribution. The Pareto becomes a shifted generalized Pareto distribution. The logistic becomes a generalized logistic distribution. The negative exponential becomes the negative exponentiated exponential, see Nadarajah and Kotz (2006).

We note in passing the well-known fact that a transformation of the variable $Y \sim \mathrm{XD}(\mu, \lambda)$ to the cumulated hazard scale $H(Y/\lambda + h^{-1}(\mu))$ gives an exponential variable with parameter λ, possibly right censored at the point $H(b)$.

4.3 Frailty models

The study of frailty models reveals a certain intimate connection between extreme and exponential dispersion models. Let the conditional distribution $Y|X = x$ be exponential with parameter x, and let X be a non-negative random variable with moment generating function $M(t)$. Then the marginal survival function for Y is $M(-y)$ for $y > 0$, which is in effect the *frailty model* of Vaupel et al. (1979).

In the special case where $X \sim \mathrm{ED}(\mu, \lambda)$ with moment generating function (17) we obtain the following survival function for Y (Hougaard, 1986):

$$G(y) = \frac{M^{\lambda}(-y/\lambda + \tau^{-1}(\mu))}{M^{\lambda}(\tau^{-1}(\mu))}.$$

This is an extreme dispersion model with hazard function $h(y) = \tau(-y)$, and corresponding unit slope function $v(\mu) = -V(\mu)$, the negative of the unit variance function for Y. This model is hence DFR. An example is the generalized Pareto distribution, which is the frailty model corresponding to a gamma frailty, with slope function $-\mu^2$ for $\mu \in (0, 1)$. As this example illustrates, the domain for v may be a proper subset of that for V.

In the particular case where X has a positive probability at 0, the distribution of Y becomes improper with $P(Y = \infty) = P(X = 0) > 0$. When X follows the Tweedie compound Poisson distribution with $1 < p < 2$ (cf. Jørgensen, 1997, Ch. 4), which has a positive probability at 0, an improper distribution for Y is obtained, as pointed out by Aalen (1988).

4.4 Exponential convergence

We shall now return to the exponential convergence of Section 2.3. By way of motivation, note that an exponential dispersion variable $Y \sim \mathrm{ED}(\mu, \lambda)$ convergences in probability to μ as $\lambda \to \infty$, as is clear from (16). The analogous result for extreme dispersion models involves convergence to the exponential distribution.

Proposition 4.1 *For $Y \sim \mathrm{XD}(\mu, \lambda)$ and $c \in \mathbb{R}$ the conditional distribution of $Y - c$ given $Y > c$ is asymptotically exponential with rate μ for $\lambda \to \infty$.*

Proof: Let λ be large enough to make the support $\lambda\left(\mathcal{C} - h^{-1}(\mu)\right)$ contain c. Then the conditional survival function of $Y - c$ given $Y > c$ is, for $y > 0$,

$$G_c(y; \mu, \lambda) = \frac{G^{\lambda}\left(\left(c + y\right)/\lambda + h^{-1}(\mu)\right)}{G^{\lambda}(c/\lambda + h^{-1}(\mu))}.$$

A Taylor expansion of H around $c/\lambda + h^{-1}(\mu)$ gives

$$G_c(y; \mu, \lambda) = \exp\left[-yh\left\{c/\lambda + h^{-1}(\mu)\right\} - \frac{y^2}{2\lambda} h'\left\{c_{\lambda\, y} + h^{-1}(\mu)\right\}\right],$$

where $c_{\lambda\, y}$ is between c/λ and $\left(c + y\right)/\lambda$. Letting $\lambda \to \infty$ and using the continuity of h and h', we obtain the desired result. ∎

5 Quadratic slope functions

We now follow Morris' (1982) footsteps and classify the set of quadratic slope functions. To this end, we need to study reflections of slope functions, and the role of exponential components. These transformations have a somewhat formal nature, but turn out to be useful for the classification result. For the sake of brevity, certain details in this section are left to the reader.

Table 2: Some vertically reflected quadratic hazard slopes.

HL(μ)	$G(y)$	$\mathcal{C}$	$v(\mu)$	Ψ
Reflected Gumbel	$\exp(1-y-e^{-y})$	$\mathbb{R}_+$	$1-\mu$	$(0,1)$
Reflected Logistic	$1/\cosh\frac{1}{2}y$	$\mathbb{R}_+$	$\left(\frac{1}{2}-\mu\right)\left(\frac{1}{2}+\mu\right)$	$(0,\frac{1}{2})$
Reflected neg. exponential	$4\left(e^{-y}-e^{-2y}\right)$	$(\log 2,\infty)$	$(\mu-1)(\mu-2)$	$(0,1)$

5.1 Reflections

We now consider what happens when we subject v to a horizontal or vertical reflection.

Proposition 5.1 *Let $G \in \overline{\mathcal{G}}_0$ have support $\mathcal{C}=(a,b)$ and slope function v.* Horizontal reflection: *If G is right censored, then the survival function $y \mapsto G(b)/G(-y)$ with support $-\mathcal{C}$ has hazard function $h(-y)$ and slope function $-v$ on Ψ, and is also right censored.* Vertical reflection: *Assume that $(0,m) \subseteq \Psi$ and restrict the support to the interval (a_0,b_0), either $(a,h^{-1}(m))$ (IFR case) or $(h^{-1}(m),b)$ (DFR case). If G is right censored at $b_0<\infty$ then the survival function $y \mapsto G(-y)/G(b_0)\exp\{-m(y+b_0)\}$ with support $(-b_0,-a_0)$ has slope function $\mu \mapsto v(m-\mu)$ with domain $(0,m)$, and is right censored if $a_0>-\infty$.*

Proof: Horizontal reflection: The survival function $G(b)/G(-y)$ with support $-\mathcal{C}$ is easily seen to have hazard function $h(-y)$ and slope function $-v(\mu)$ on Ψ. The value at the right endpoint is $G(b)/G(a)>0$, so the model is right censored. Vertical reflection: The survival function $G(-y)/G(b_0)\exp\{-m(y+b_0)\}$ with support $(-b_0,-a_0)$ is similarly seen to have hazard function $m-h(-y)$ and slope function $v(m-\mu)$ on $(0,m)$. If $a_0>-\infty$ then $G(a_0)/G(b_0)\exp\{-m(b_0-a_0)\}>0$, so in this case the model is right censored. ■

Table 2 shows three hazard location families with quadratic slope functions obtained by vertical reflection of families from Table 1.

5.2 Exponential components

Extending the results of Section 2.3, we now show that an exponential component (in the sense of Remark 5.1 below) corresponds to a location change for the slope function.

Proposition 5.2 *Let $G \in \overline{\mathcal{G}}_0$ with support $\mathcal{C}=(a,b)$ have slope function v with domain $\Psi=(\underline{\eta},\overline{\eta})$. If $a>-\infty$ then for $m \geq -\underline{\eta}$, the function $v(\mu-m)$ with domain $m+\Psi$ is the slope function of the survival function given by $G(y)\exp\{-m(y-a)\}$ on $\mathcal{C}$.*

Proof: The survival function $G(y)\exp\{-m(y-a)\}$ has integrated hazard function

$$m(y-a)+H(y) \tag{21}$$

on $\mathcal{C}$, provided $m \geq -\underline{\eta}$, with hazard function $m+h(y)$ and slope function $v(\mu-m)$ on $m+\Psi$. ■

Remark 5.1 *A positive m in (21) corresponds to the variable* $\min\{Y,a+E_m\}$, *where the exponential variable E_m is independent of Y. In this case we say that we are* introducing *an exponential component. Conversely, when $m=-\underline{\eta}<0$ we say that we are* removing *the exponential component, making* $\inf\Psi=0$. *The only model in Table 1 with an exponential component is the uniform distribution. After removing the exponential component, we obtain*

$$G(y)=e^y(1-y) \text{ for } y\in(0,1). \tag{22}$$

The corresponding slope function is $(1+\mu)^2$ with domain $\Psi=\mathbb{R}_+$.

Note that, using the terminology of Remark 5.1, it is understood in connection with the classification results below that vertical reflection (Proposition 5.1) is applied only after removing the exponential component, if necessary, to ensure that $\inf \Psi = 0$.

An example of an exponential component is encountered in connection with the Gumbel family with unit slope function $v(\mu) = \mu$ on $\Psi = \mathbb{R}_+$. The Gompertz-Makeham distribution is obtained from the Gumbel by left truncation at 0, restricting v to $\mu > 1$, and then adding an exponential component. This gives the hazard function $h(y) = m + e^{\beta y}$ for $m, \beta, y > 0$ and unit slope function $v(\mu) = \mu - m$ for $\mu > 1 + m$. A horizontal reflection, corresponding to $\beta < 0$, yields $v(\mu) = m - \mu$ for $\mu \in (m, m+1)$.

5.3 Classification

We have now considered several types of transformations of slope functions, including censoring, truncation and reflections. In addition to these, we consider the following three transformations of a given hazard location family $\mathrm{HL}(\mu)$ with slope function v and domain Ψ.

1. Location change: removing or adding an exponential component maps v into $v(\mu - m)$.
2. Scale transformation: a scale transformation of Y maps v into $c^{-2}v(c\mu)$ for $c > 0$.
3. Multiplication: generating an extreme dispersion model maps v into v/λ for $\lambda > 0$.

A combination of these three operations maps v into

$$\gamma v\left(\left(\mu - \alpha\right)/\beta\right), \tag{23}$$

where $\gamma, \beta > 0$ and $\mu \in \alpha + \beta\Psi$. We refer to (23) as the operation of *location and scaling*. This leads us to the main classification theorem.

Theorem 5.3 *Up to left truncation, right censoring, reflection, location and scaling, the only hazard location models with quadratic slope functions are those shown in Table 1.*

The next remark will useful for the proof.

Remark 5.2 *Consider $G \in \overline{\mathcal{G}}$ and $c, d \in \mathcal{C}$. The following identity*

$$H(d) - H(c) = \int_{h(c)}^{h(d)} \frac{\mu}{v(\mu)}\, d\mu \tag{24}$$

follows by the substitution $\mu = h(x)$ in the integral (1). It is useful for checking if a given function v may serve as a slope function. By taking $c = a$, we find that the continuity of G at a is equivalent to the integral (24) being convergent at $h(a)$. By taking $d = b$, we find that right censoring is equivalent to the integral (24) being convergent at $h(b)$.

Proof: [of Theorem 5.3] By means of the location and scaling operation we may reduce the classification problem to quadratic slope functions with simple forms like in Table 1, having roots either ± 1, 0 or i. A combination of vertical and horizontal reflections applied to the seven cases of Table 1 then covers all possible shapes of quadratic slope functions, most of which are right censored (cf. Proposition 5.1). Regarding the uniform and Pareto distributions, an application of Remark 5.2 shows that neither μ^2 nor $-\mu^2$ can be slope functions on $\Psi = \mathbb{R}_+$, but only on a subset of $\mathbb{R}_+$. It follows that horizontal reflections of the uniform and Pareto distributions give rise to two separate cases of right censored slope functions of the form $\pm\mu^2$, and a further four cases of the form $(\mu - m)^2$ that follow by vertical reflection. It is easily seen that this covers all possible cases. ∎

Table 3: Summary of generalized extreme value distributions.

$\mathrm{EV}_\gamma(\mu,\lambda)$	γ	p	Support
Weibull	$\gamma < -1$	$p > 2$	$(1/\gamma, \infty)$
Exponential	$\gamma = -1$	$-$	$(-1, \infty)$
Weibull	$-1 < \gamma < 0$	$p < 1$	$(1/\gamma, \infty)$
Gumbel	$\gamma = 0$	$p = 1$	$\mathbb{R}$
Fréchet	$\gamma > 0$	$1 < p < 2$	$(-\infty, 1/\gamma)$

Remark 5.3 *There are three cases with* $\inf \mathcal{C} = -\infty$ *in Table 1 where a vertical reflection leads to models that are not right censored, of which typical examples are shown in Table 2.*

One could also explore parallels of other classification results for NEFs, such as Letac and Mora's (1990) cubic variance functions, but this is outside the scope of the present paper.

6 Generalized extreme value distributions

We now investigate the analogy between the generalized extreme value distribution and the Tweedie class of exponential dispersion models. The latter is characterized by having unit variance functions of power form $V(\mu) = \mu^p$, cf. Jørgensen (1997, Ch. 4) and references therein. Here $p = 0$ corresponds to the normal distribution with domain $\mathbb{R}$, whereas the remaining cases, namely $p < 0$ and $p \geq 1$, all have domain $\mathbb{R}_+$. The special cases $p = 0, 1, 2$ all appear in Table 1, and a further simple case is $p = 3$, corresponding to the inverse Gaussian distribution.

6.1 Definition

The standard generalized extreme value distribution for minima is defined by

$$G(y) = \exp\left\{-\left(1-\gamma y\right)^{-1/\gamma}\right\},$$

with support defined by $\gamma y < 1$. Here $\gamma \in \mathbb{R}$, and the value $\gamma = 0$ (defined by continuity) corresponds the Gumbel distribution. All extreme value distributions except the exponential ($\gamma = -1$) have monotone hazard rates and their slope functions are of power form

$$v(\mu) = \frac{1}{2-p}\mu^p \tag{25}$$

for $\mu > 0$, where the parameter $p \in \mathbb{R}\setminus\{2\}$ is defined by

$$p = p(\gamma) = \frac{1+2\gamma}{1+\gamma}. \tag{26}$$

The models are IFR for $p < 2$ ($\gamma > -1$) and DFR for $p > 2$ ($\gamma < -1$). As we saw in the proof of Theorem 5.3 there is no slope function on $\mathbb{R}_+$ proportional to μ^2. Table 3 summarizes the main cases of generalized extreme value distributions corresponding to different values of p.

Introducing location and index parameters, the generalized extreme value distributions are seen to be examples of extreme dispersion models, one for each γ. We thus define the $\mathrm{EV}_\gamma(\mu,\lambda)$ to be the extreme dispersion model given by the survival function

$$y \mapsto \exp\left\{-\lambda\left(\mu^{-\gamma/(1+\gamma)} - \gamma y/\lambda\right)^{-1/\gamma}\right\},$$

with support defined by $\gamma y < \lambda\mu^{-\gamma/(1+\gamma)}$, which is a reparametrization of the usual extreme value distribution. The $\mathrm{EV}_\gamma(\mu,\lambda)$ model satisfies the following scaling property

$$c\mathrm{EV}_\gamma(\mu,\lambda) = \mathrm{EV}_\gamma(c^{-1}\mu, c^{2-p}\lambda). \tag{27}$$

6.2 Characterization

The Tweedie models may be characterized as the only exponential dispersion models closed under scale transformations, cf. Jørgensen (1997, p. 128). We now show, by means of Proposition 3.1 that the extreme value distributions satisfy a similar property.

Theorem 6.1 *Let $XD(\mu,\lambda)$ be such that for some $\lambda > 0$ and all $\mu, c > 0$*

$$c\mathrm{XD}(\mu,\lambda) = \mathrm{XD}(c^{-1}\mu, g_\lambda(c)) \tag{28}$$

for some positive function $g_\lambda(c)$. Then $XD(\mu,\lambda)$ is a generalized extreme value distribution.

Proof: First note that the rate $c^{-1}\mu$ on the right-hand side of (28) is consistent with (4). Since $c > 0$ is arbitrary this in turn implies that $\Psi = \mathbb{R}_+$. Without loss of generality we may take $\lambda = 1$. Calculating the slope function on both sides of (28) gives

$$c^{-2}v(\mu) = \frac{1}{g_1(c)}v(c^{-1}\mu).$$

This implies that v satisfies the functional equation $v(x)v(y) = v(1)v(xy)$ for $x, y > 0$. Using the continuity of v, the solution is $v(\mu) = c_p\mu^p$, where $p \in \mathbb{R}$ and c_p is an arbitrary non-zero constant that may depend on p. When $p \neq 2$ and $c_p = 1/(2-p)$ this characterizes the generalized extreme value distribution $\mathrm{EV}_\gamma(\mu,\lambda)$. Other choices for c_p with the same sign correspond to a scale change. Changing the sign to $c_p = -1/(2-p)$ is possible only for a right censored survival function (Proposition 5.1), which is incompatible with the condition $\Psi = \mathbb{R}_+$. The case $p = 2$, which has been dealt with in Section 5, also is not compatible with the condition $\Psi = \mathbb{R}_+$. It easily follows that $g_\lambda(c) = \lambda c^{2-p}$, in agreement with (27). ■

7 Convergence of extremes

7.1 General convergence theorem

The results of the previous section show that the Tweedie and generalized extreme value distributions share certain properties due to the common form of their variance and slope functions. We shall now complete this analogy by showing a convergence theorem for slope functions, which in turn leads to a new proof of the extreme value convergence theorem, along the same lines as the Tweedie convergence theorem of Jørgensen et al. (1994).

The use of variance functions for proving convergence for natural exponential families was initiated by Morris (1982), but a rigorous formulation and proof was first given by Mora (1990). The convergence theorem for variance functions says that if a sequence of variance functions converges uniformly on compact sets, then the corresponding sequence of natural exponential families converges to the family corresponding to the limiting variance function. We have the following analogous result for slope functions. The proof is given in an Appendix.

Theorem 7.1 *Let v_n, $\Psi_n = (\underline{\eta}_n, \overline{\eta}_n)$ be a sequence of slope functions and their respective domains, all IFR (DFR), such that $\Psi = \operatorname{int}(\lim_{n\to\infty} \Psi_n)$ exists and is non-empty, where $\lim_{n\to\infty} \Psi_n$ means that each of the two sequences of endpoints converges. Assume that v_n converges on Ψ, uniformly on compact subintervals of Ψ, to a function v which is strictly positive (strictly negative) on Ψ. Assume that v_n satisfies the following left tightness condition. For each $k > 0$ there exists an $\eta \in \Psi$ such that for all n*

$$\int_{I_n(\eta)} \frac{\mu}{|v_n(\mu)|}\, d\mu < k, \tag{29}$$

where $I_n(\eta) = (\underline{\eta}_n, \eta)$ ($I_n(\eta) = (\eta, \overline{\eta}_n)$). Then the corresponding sequence of hazard location families $\mathrm{HL}_n(\mu)$ converges weakly for each $\mu \in \Psi$, uniformly on compact subintervals of the support, to the hazard location family $\mathrm{HL}(\mu)$ with slope function v.

The tightness condition (29) originates from the identity (24). The following remark shows that a similar condition is useful for determining if the limiting family is right censored.

Remark 7.1 *Under the assumptions of Theorem 7.1, we consider the following right tightness condition. For each $k > 0$ there exists an $\eta \in \Psi$ such that for all n*

$$\int_{I_n(\eta)} \frac{\mu}{|v_n(\mu)|}\, d\mu > k,$$

where $I_n(\eta) = (\eta, \overline{\eta}_n)$ (IFR case) or $I_n(\eta) = (\underline{\eta}_n, \eta)$ (DFR case). Then for every $k > 0$ there exists a $c \in \mathcal{C}$ such that $H_n(c) > k$ for all n. This, in turn, implies that $H(c) = \lim H_n(c) > k$, and hence $h(b-) = \infty$. This implies no right censoring, so in particular the limiting distribution is proper.

7.2 Extreme convergence theorem

Let $\gamma \neq -1$ be given, and let $p = p(\gamma) \neq 2$, according to (26). Choosing c in the scaling formula (27) such that $n = c^{2-p}$ is an integer, we obtain

$$n^{1/(p-2)}\mathrm{EV}_\gamma(n^{1/(p-2)}\mu, n\lambda) = \mathrm{EV}_\gamma(\mu, \lambda). \tag{30}$$

Now recall the min reproductive property (20), by which the left-hand side of (30) represents a centering and scaling of the scaled min $\hat{Y}_n$ for a sample of size n from $\mathrm{EV}_\gamma(\mu, \lambda)$. In effect (30) represents the so-called *stability postulate* for the limiting distribution of extremes, cf. Kotz and Nadarajah (2002, p. 5). The corresponding domains of attraction correspond to the classical extreme convergence result, which in the present setup takes the following form.

Theorem 7.2 *Let $\mathrm{XD}(\mu, \lambda)$ be an extreme dispersion model having unit slope function v with power asymptotics of the form*

$$v(\mu) \sim \frac{1}{2-p}\mu^p \tag{31}$$

as $\mu \to 0$ (IFR case with $p < 2$) or $\mu \to \infty$ (DFR case with $p > 2$). Then for any $\mu, \lambda > 0$

$$n^{1/(p-2)}\mathrm{XD}(n^{1/(p-2)}\mu, n\lambda) \xrightarrow{w} \mathrm{EV}_\gamma(\mu, \lambda) \text{ as } n \to \infty, \tag{32}$$

where $\xrightarrow{w}$ denotes weak convergence.

Proof: Consider the IFR case $p < 2$, where the power asymptotics holds near 0. For fixed values of λ and n, the left-hand side of (32) is a hazard location family with slope function

$$v_n(\mu) = \frac{1}{\lambda n^{p/(p-2)}} v(n^{1/(p-2)}\mu) \longrightarrow \frac{1}{\lambda\,(2-p)}\mu^p \text{ as } n \to \infty,$$

where we have used the scaling property of the slope in (4). The pointwise convergence follows from (31). To show that the convergence is uniform in μ on compact subsets of $\mathbb{R}_+$, let $0 < \mu < m$ for given $m > 0$. For given $\varepsilon > 0$ let μ_0 be such that

$$\left| \frac{v(\mu)}{\mu^p} - \frac{1}{2-p} \right| < \varepsilon$$

for $\mu < \mu_0$, by the assumption of power asymptotics. Then for any n large enough to make $n^{1/(p-2)} < \mu_0/m$ we find

$$\left| \frac{v(n^{1/(p-2)}\mu)}{n^{p/(p-2)}} - \frac{\mu^p}{2-p} \right| = \mu^p \left| \frac{v(n^{1/(p-2)}\mu)}{\left(\mu n^{1/(p-2)}\right)^p} - \frac{1}{2-p} \right| \leq m^p \varepsilon$$

for all $\mu < m$, which shows the uniform convergence. Since we are in the IFR case, the tightness condition (29) involves the integral

$$\int_0^\eta \frac{\lambda n^{p/(p-2)}\mu}{v(n^{1/(p-2)}\mu)}\, d\mu.$$

Since the integrand behaves asymptotically like the power μ^{1-p}, which is integrable on $(0, \eta)$ for $p < 2$, the tightness condition is satisfied. The convergence (32) hence follows from Theorem 7.1. The proof in the DFR case $p > 2$ is similar. ■

Compared with conventional extreme value results, the framework of Theorem 7.2 is very convenient, albeit under the rather strong conditions of differentiability of the density, and monotone hazard rate. We note that the condition (31) seamlessly integrates the Gumbel case ($p = 1$) with the rest, whereas the exponential case is not included, see Remark 7.3.

We note that the approach leads to the new centering constant $\theta = -h^{-1}(n^{1/(p-2)}\mu)$, the location parameter appearing in (18), which corresponds to keeping the rate constant at the value μ throughout the convergence (32).

In simple cases, like in Table 1, it is very easy to read off the asymptotic behaviour of the slope function. For example, the asymptotic behaviour of v near 0 for the logistic and negative exponential distributions is $v(\mu) \sim \mu$, so both are in the domain of attraction of the Gumbel distribution. Further examples are considered below.

Remark 7.2 *The von Mises conditions are sufficient conditions involving the density f for extreme value convergence. For $G \in \overline{\mathcal{G}}$ with support (a, b), the version of the Gumbel condition proposed by Falk and Marohn (1993) is (keeping in mind that we use min rather than max)*

$$\lim_{y \downarrow a} \frac{f(y)}{1 - G(y)} = c$$

for some $c > 0$. By l'Hospital's rule this is equivalent to

$$\lim_{y \downarrow a} \frac{h'(y)}{h(y)} = c. \tag{33}$$

By inserting $y = h^{-1}(\mu)$, we find that (33) is equivalent to (31) with $p = 1$. The situation for $p \neq 1$ is, however, less clear. For $a = 0$ the Gumbel condition is

$$\lim_{y \downarrow 0} \frac{y f(y)}{1 - G(y)} = -\gamma^{-1} > 0,$$

or equivalently, with an application of l'Hospital's rule,

$$\lim_{y \downarrow 0} \frac{y h'(y)}{h(y)} = -1 - \gamma^{-1}.$$

This condition apparently cannot be expressed conveniently in terms of the slope function v.

Remark 7.3 *Contrary to conventional extreme value convergence theory, our framework separates out the case of exponential convergence, and Proposition 4.1 illustrates how exponential convergence is prompted by left truncation, see also (9). The uniform and Pareto examples from Table 1 illustrate that distributions in the domain of attraction of the exponential distribution have incomplete* Ψ*, with* $\inf \Psi > 0$ *(IFR case) or* $\sup \Psi < \infty$ *(DFR case). In the case of the uniform distribution with the exponential component removed (22), the new slope function* $(1 + \mu)^2$ *satisfies (31) with* $p = 0$*, and so is in the domain of attraction of the Rayleigh distribution.*

Jørgensen and Martínez (1997) developed Tauberian methods for variance functions, where power asymptotics for V is replaced by regular variation. This could be developed in the present setting along the lines of de Haan (1970), but is outside the scope of the present paper.

7.3 Examples

Let us consider two further examples of extreme value convergence that illustrate Theorem 7.2. First we consider the *negative Pareto distribution* with survival function $G(y) = 1 - (1 - y)^{-1}$ for $y < 0$. Straightforward calculations show that the corresponding slope function is

$$v(\mu) = \mu\sqrt{\mu^2 + 4\mu} \text{ for } \mu > 0, \tag{34}$$

which behaves like $2\mu^{3/2}$ near 0. Letting $\mathrm{XD}(\mu, \lambda)$ denote the extreme dispersion model corresponding to G, an application of Theorem 7.2 yields Fréchet convergence,

$$n^{-2}\mathrm{XD}(n^{-2}\mu, n\lambda) \xrightarrow{w} \mathrm{EV}_1(\mu, \lambda) \text{ as } n \to \infty.$$

It is worth noting that v in (34) is of the so-called *Letac form* (Jørgensen, 1997, pp. 157–158), a class of variance functions that has been extensively studied, see e.g. Kokonendji (1994).

Next, we consider the *Burr distribution* with survival function $G(y) = (1 + y^\alpha)^{-1}$ for $y > 0$, for some $\alpha > 0$, which is DFR for $0 < \alpha \leq 1$. An explicit expression for the slope function may be found in the case $\alpha = 1/2$, where

$$v(\mu) = -\mu^2 \left(\mu + 2 + \sqrt{\mu^2 + 2\mu}\right) \text{ for } \mu > 0.$$

The asymptotic behaviour is $v(\mu) \sim -2\mu^3$ as $\mu \to \infty$. An application of Theorem 7.2 yields Weibull convergence with $\gamma = -2$,

$$n\mathrm{XD}(n\mu, n\lambda) \xrightarrow{w} \mathrm{EV}_{-2}(\mu, \lambda/2) \text{ as } n \to \infty,$$

where $\mathrm{XD}(\mu, \lambda)$ denotes the extreme dispersion model generated by G. For $0 < \alpha < 1$ the behaviour of v is like $-\mu^p$ with $p = (\alpha - 2)/(\alpha - 1) > 2$, and (32) applies.

For $\alpha > 1$ the Burr hazard is not monotone, but is for y near 0. Hence by a suitable right censoring, we obtain an IFR model with asymptotic behaviour μ^p for v with $p < 1$. In general Theorem 7.2 may be applied in this way to models with non-monotone hazard as long as the hazard is monotone near 0.

8 Exponential slope functions

We now consider characterization and convergence for exponential slope functions, similar to Jørgensen's (1997, p. 160) characterization of exponential variance functions. These results have independent interest, since exponential variance functions correspond to natural exponential families generated by extreme stable distributions with stability index $\alpha = 1$.

8.1 Characterization

Elaborating on the parallel between the Rayleigh and normal distributions, we note that the latter satisfies the following transformation property:

$$\mathrm{N}(m+\mu,\sigma^2) - m = \mathrm{N}(\mu,\sigma^2)$$

for all $m \in \mathbb{R}$, imitating (30), but with multiplication replaced by addition. From (10), we would expect in the Rayleigh IFR case that the term $-m$ corresponds to a left truncation followed by the removal of an exponential component. More generally, given $\mathrm{XD}(\mu,\lambda)$ with unit slope function v on $\psi = \mathbb{R}_+$, we consider the *shift transformation*, defined by the following two steps.

1. Left truncation (IFR) or right censoring (DFR), which restricts the domain to $\mu > m$, while maintaining the slope at $\lambda^{-1}v(\mu)$. This gives rise to an exponential component.

2. Removing the exponential component, giving the rate $\mu > 0$ and slope $\lambda^{-1}v(m+\mu)$.

The result is an extreme dispersion model $\mathrm{XD}_m(\mu,\lambda)$ with unit slope function $v(m+\cdot)$. We now characterize exponential slope functions as fixed points for the shift transformation.

Theorem 8.1 *Let $XD(\mu,\lambda)$ have unit slope function v and domain $\Psi = \mathbb{R}_+$. If for some $\lambda > 0$ there exists a positive function $g_\lambda(m)$ such that for all $m,\mu > 0$*

$$\mathrm{XD}_m(\mu, g_\lambda(m)) = \mathrm{XD}(\mu,\lambda), \tag{35}$$

then the unit slope function v is either constant or exponential.

Proof: By calculating the slope on both sides of (35) we obtain the equation

$$\frac{1}{g_\lambda(m)} v(m+\mu) = \lambda^{-1}v(\mu).$$

Without loss of generality we may take $\lambda = 1$. By letting $\mu \downarrow 0$ and using the continuity of v, we find that the limit $v(0+)$ exists, is positive and finite, and $g_1(m) = v(m)/v(0+)$. This, in turn, implies that v satisfies the functional equation $v(0+)v(m+\mu) = v(m)v(\mu)$ for all $m,\mu > 0$. Taking into account the continuity of v, the solution is

$$v(\mu) = v(0+)e^{\beta\mu} \tag{36}$$

for some $\beta \in \mathbb{R}$, which in turn implies that (35) holds for all $\lambda > 0$ with $g_\lambda(m) = \lambda e^{\beta m}$. ■

Besides the Rayleigh case ($\beta = 0$) there are two main cases of (36), one IFR and one DFR. The IFR case has unit slope function $v(\mu) = e^{-\mu}$ for $\mu > 0$, and corresponds to the extreme dispersion model generated from the survival function

$$G(y) = e^y\,(1+y)^{-(1+y)} \text{ for } y > 0. \tag{37}$$

The DFR case has unit slope function $v(\mu) = -e^{\mu}$ for $\mu > 0$, and corresponds to the extreme dispersion model generated from the survival function

$$G(y) = e^{-y}y^y \text{ for } 0 < y < 1, \tag{38}$$

which is right censored at 1.

Note that by applying a suitable location and scaling operation to the power slope function (25) for $p > 2$ we obtain

$$-\left(1+\frac{\mu}{p}\right)^p \to -e^{\mu} \text{ for } p \to \infty,$$

which shows that the DFR case of (36) is a limiting case of the generalized extreme value family. A similar result holds in the IFR case.

8.2 Convergence

We now show a convergence theorem for exponential slope functions, similar to a result for exponential variance functions (Jørgensen, 1997, p. 164). In effect, the fixed point (35) has a domain of attraction consisting of models with asymptotically exponential slope functions.

Theorem 8.2 *Let* $\mathrm{XD}(\mu, \lambda)$ *denote an extreme dispersion model with unit slope function* v *and domain* $\Psi = \mathbb{R}_+$ *having exponential asymptotics of the form*

$$v(\mu) \sim c_\beta e^{\beta\mu} \tag{39}$$

for $\mu \to \infty$*, where* $c_\beta = 1$ *for* $\beta \leq 0$ *and* $c_\beta = -1$ *for* $\beta > 0$*. Then the shifted model* $XD_m(\mu, \lambda e^{\beta m})$ *converges to an extreme dispersion model with exponential slope function for* $m \to \infty$*.*

Proof: The shifted model $\mathrm{XD}_m(\mu, \lambda e^{\beta m})$ has unit slope function

$$e^{-\beta m} v(m + \mu) \to c_\beta e^{\beta\mu} \text{ for } m \to \infty, \tag{40}$$

pointwise for $\mu > 0$. To show that the convergence is uniform in μ on compact subsets of $\mathbb{R}_+$, let $0 < \mu < m_0$ for given $m_0 > 0$. For given $\varepsilon > 0$ let μ_0 be such that

$$\left| e^{-\beta\mu} v(\mu) - c_\beta \right| < \varepsilon$$

for $\mu > \mu_0$, by (39). Then for any $m > \mu_0$ we find

$$\left| e^{-\beta m} v(m + \mu) - c_\beta e^{\beta\mu} \right| = e^{\beta\mu} \left| e^{-\beta(m+\mu)} v(m + \mu) - c_\beta \right| \leq \left(1 + e^{\beta m}\right) \varepsilon$$

for all $\mu < m_0$, showing uniform convergence. The tightness condition (29) involves the integral

$$\int_{I_m(\eta)} \frac{\lambda\mu}{|e^{-\beta m} v(m + \mu)|} \, d\mu,$$

where the integrand behaves asymptotically like $\mu e^{-\beta\mu}$. For $\beta > 0$ the interval of integration is (η, ∞), whereas for $\beta \leq 0$ it is $(0, \eta)$, so in both cases the tightness condition is satisfied. The result now follows from Theorem 7.1. ∎

There are three main cases of (40). *DFR case.* Take $\beta = 1$ and let m be such that $n = e^m$ is an integer. We may then write the convergence as follows:

$$\mathrm{XD}_{\log n}(\mu, \lambda n) \xrightarrow{w} \mathrm{XD}_-(\mu, \lambda) \text{ for } n \to \infty, \tag{41}$$

where $\mathrm{XD}_-(\mu, \lambda)$ is the model generated by (38). The left-hand side of (41) represents a shift transformation of the scaled min $\hat{Y}_n$ for a sample of size n from $\mathrm{XD}(\mu, \lambda)$. *Rayleigh case.* For $\beta = 0$ we obtain convergence to the Rayleigh distribution,

$$\mathrm{XD}_m(\mu, \lambda) \xrightarrow{w} \mathrm{EV}_{-\frac{1}{2}}(\mu, \lambda) \text{ for } m \to \infty.$$

IFR case. Take $\beta = -1$ and let $t = e^{-m}$. Then

$$\mathrm{XD}_{-\log t}(\mu, \lambda t) \xrightarrow{w} \mathrm{XD}_+(\mu, \lambda) \text{ for } t \downarrow 0,$$

where $\mathrm{XD}_+(\mu, \lambda)$ is the model generated by (37). This in effect involves the asymptotic distribution of an extremal process X_t for $t \downarrow 0$, much like the infinitely divisible type of convergence of Jørgensen (1997, p. 149). This follows by noting that to every $\mathrm{XD}(\mu, \lambda)$ model there exists an extremal process X_t, in the sense of Dwass (1964), such that $tX_t \sim \mathrm{XD}(\mu, \lambda t)$.

Appendix: Proof of general convergence theorem

The proof of Theorem 7.1 proceeds along the same lines as Jørgensen's (1997, p. 54) proof of the convergence theorem for variance functions, which in turn is a simplification of Mora's (1990) proof in the multivariate case. The idea is to reconstruct the hazard function h from the limiting slope function v using (14), and in turn use the uniform convergence and tightness to show convergence of the sequence H_n.

Let K be a given compact subinterval of Ψ. By assumption $\Psi = \operatorname{int}(\lim \Psi_n)$, so we may assume that $K \subseteq \Psi_n$ from some n_0 on. We only need to consider $n > n_0$. Fix a $\mu_0 \in \operatorname{int} K$. Let $\psi_n = h_n^{-1}$ denote the inverse hazard function given by $\psi_n'(\mu) = 1/v_n(\mu)$ on Ψ_n and $\psi_n(\mu_0) = 0$, cf. (14). Let h_n, H_n etc. denote the quantities associated with this parametrization.

Similarly, define $\psi : \Psi \to \mathbb{R}$ by $\psi'(\mu) = 1/v(\mu)$ on Ψ and $\psi(\mu_0) = 0$. Then for $\mu \in K$

$$|\psi_n'(\mu) - \psi'(\mu)| = \frac{|v_n(\mu) - v(\mu)|}{v_n(\mu)v(\mu)}. \tag{42}$$

By the uniform convergence of $v_n(\mu)$ to $v(\mu)$ on K, it follows that $v_n(\mu)$ is uniformly bounded on K. Since $v(\mu)$ is bounded on K, it follows from (42) and from the uniform convergence of v_n that $\psi_n'(\mu) \to \psi'(\mu)$ uniformly on K. This and the fact that $\psi_n(\mu_0) = \psi(\mu_0)$ for all n implies, by a result from Rudin (1976, Theorem 7.17, p. 152), that $\psi_n(\mu) \to \psi(\mu)$ uniformly on K.

Let $\mathcal{C}_n = \psi_n(\Psi_n)$ and $\mathcal{C} = \psi(\Psi)$. Then $\mathcal{C} = \operatorname{int}(\lim \mathcal{C}_n)$. Let $J = \psi(K) \subseteq \mathcal{C}$ and $J_n = \psi_n(K) \subseteq \mathcal{C}_n$. Define $h : \mathcal{C} \to \Psi$ by $h(y) = \psi^{-1}(y)$. Since ψ is strictly monotone and differentiable, the same is the case for h, and $h(y) > 0$ on $\mathcal{C}$ since $\Psi \subseteq \mathbb{R}_+$. Let $\mu \in K$ be given and let $y = \psi(\mu) \in J$ and $y_n = \psi_n(\mu) \in J_n$. Since $v_n(\mu)$ is uniformly bounded on K, there exists an m such that $|v_n(\mu)| \leq m$ for all n and $\mu \in K$. It follows that $|h_n'(y)| \leq m$ for all $y \in J$. Since $\mu = h(y) = h_n(y_n)$ we find, using the mean value theorem, that

$$\begin{aligned} |h_n(y) - h(y)| &= |h_n(y) - h_n(y_n)| \\ &\leq m\,|y - y_n| \\ &= m\,|\psi(\mu) - \psi_n(\mu)|\,. \end{aligned}$$

This implies that $h_n(y) \to h(y)$ uniformly in $y \in J$. The above arguments also apply if J is extended to a larger subinterval of $\mathcal{C}$.

In order to invoke the tightness condition (29), we first consider the IFR case. Using (24) we obtain for $c \in \mathcal{C}$

$$H_n(c) = \int_{\underline{\eta}_n}^{h_n(c)} \frac{\mu}{|v_n(\mu)|}\, d\mu. \tag{43}$$

For a given $\eta \in \Psi$ we may choose $\varepsilon > 0$ and $c \in \mathcal{C}$ such that $h(c) + \varepsilon < \eta$, and from the convergence of $h_n(c)$ to $h(c)$ we obtain $\underline{\eta}_n < h_n(c) < \eta$ for n large enough. Together with (29) this implies that for every $k > 0$ there exists a $c = c(k) \in \mathcal{C}$ such that for all n

$$0 \leq H_n(c) \leq k. \tag{44}$$

Since all H_n are increasing we can make $c(k)$ an increasing function of k. In the DFR case the inequality (44) follows similarly by integrating over the interval $(h_n(c), \overline{\eta}_n)$ in (43).

The condition (44) implies that there exists a $c \in \mathcal{C}$ such that

$$\lim \inf_{n\to\infty} \int_{-\infty}^{c} h_n(x)\, dx = \lim \inf_{n\to\infty} H_n(c) < \infty. \tag{45}$$

By Fatou's lemma, (45) implies that $\int_{-\infty}^{c} h(x)\,dx < \infty$. We may now define G and H for $y \in \mathbb{R}$ by $H(y) = \int_{-\infty}^{y} h(x)\,dx$, and $G(y) = \exp\{-H(y)\}$, where we use the conventions discussed in connection with (1). Then G is a survival function with support $\mathcal{C}$, and $H(\inf \mathcal{C}) = 0$. Using the above-mentioned result from Rudin once more, we find that for any given d in J, $H_n(y) - H_n(d)$ converges to $H(y) - H(d)$ uniformly in $y \in J$.

To conclude the proof, we choose a $d \in J$, and show that $H_n(d)$ converges to $H(d)$. The tightness condition (44) implies that, for given $k > 0$ and $c \leq c(k)$, $H_n(d)$ satisfies

$$H_n(d) - H_n(c) \leq H_n(d) \leq H_n(d) - H_n(c) + k.$$

We may enlarge J to include c. Letting $n \to \infty$ we find that $H_n(d)$ is asymptotically squeezed between the values $H(d) - H(c)$ and $H(d) - H(c) + k$, which can be made arbitrarily close to $H(d)$ by choosing k small, and c close to $\inf \mathcal{C}$. Hence $H_n(d)$ converges to $H(d)$. It follows that $H_n(y)$ converges to $H(y)$ uniformly in $y \in J$, completing the proof.

Acknowledgements

The research was supported by the Danish Natural Science Research Council.

References

- Aalen, O.O., Heterogeneity in survival analysis, Statist. Med. **7**, 1121–1137, (1988).
- Bar-Lev, S.K. and Enis, P., Reproducibility and natural exponential families with power variance functions, Ann. Statist. **14**, 1507–1522, (1986).
- Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J., Statistics of Extremes: Theory and Applications, Chichester: Wiley, 2004.
- Casalis, M., Natural exponential families. In Kotz, S., Balakrishnan, N., and Johnson, N.L. (ed.), Continuous Multivariate Distributions, Vol. 1: Models and Applications, John Wiley & Sons, New York, pp. 659–696, 2000.
- Coles, S.G., An Introduction to Statistical Modeling of Extreme Values, Springer-Verlag, London, 2001.
- de Haan, L., On Regular Variation and Its Application to the Weak Convergence of Sample Extremes, Mathematical Centre Tracts No. 32, Mathematisch Centrum, Amsterdam, 1970.
- Dwass, M., Extremal processes. Ann. Math. Statist. **35**, 1718–1725, (1964).
- Falk, M. and Marohn, F., von Mises conditions revisited. Ann. Probab. **21**, 1310–1328, (1993).
- Fisher, R.A. and Tippett, L.H.C., Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proc. Cambridge Phil. Soc. **24**, 180–190, (1928).
- Gnedenko, B.V., Sur la distribution limite du terme maximum d'une série aléatorire. Ann. Math. **44**, 423–453, (1943).
- Hougaard, P., Survival models for heterogeneous populations derived from stable distributions, Biometrika **73**, 387–396, (1986).

- Jørgensen, B., Exponential dispersion models (with discussion), J. Roy. Statist. Soc. Ser. B **49**, 127–162, (1987).
- Jørgensen, B., The Theory of Dispersion Models, Chapman & Hall, London, 1997.
- Jørgensen, B., and Martínez, J.R., Tauber theory for infinitely divisible variance functions, Bernoulli **3**, 213–224, (1997).
- Jørgensen, B., Martínez, J.R. and Tsao, M. Asymptotic behaviour of the variance function, Scand. J. Statist. **21**, 223–243, (1994).
- Klugman, S.A., Panjer, H.H. and Willmot, G.E., Loss Models: From Data to Decisions, 2nd ed., Wiley Interscience, Hoboken, N.J., 2004.
- Kotz, S. and Nadarajah, S., Extreme Value Distributions: Theory and Applications, Imperial College Press, London, 2002.
- Kokonendji, C.C., Exponential families with variance functions in $\sqrt{\Delta}P(\sqrt{\Delta})$: Seshadri's class, Test **3**, 123–172, (1994).
- Lauritzen, S.L., Thiele: Pioneer in Statistics, Oxford University Press, Oxford, 2002.
- Letac, G. and Mora, M., Natural real exponential families with cubic variance functions, Ann. Statist. **18**, 1–37, (1990).
- McCullagh, P. and Nelder, J.A., Generalized Linear Models, 2nd Ed., Chapman and Hall, London, 1989.
- Mora, M., Convergence of the variance functions of natural exponential families, Ann. Fac. Sci. Univ. Toulouse, Série 5 **11**, 105–120, (1990).
- Morris, C.N., Models for positive data with good convolution properties, Memo no. 8949. California: Rand Corporation, (1981).
- Morris, C.N., Natural exponential families with quadratic variance functions, Ann. Statist. **10**, 65–80, (1982).
- Nadarajah, S. and Kotz, S., The exponentiated type distributions, Acta Appl. Math. **92**, 97–111, (2006).
- Nelson, W. and Doganaksoy, N., Statistical analysis of life or strength data from specimens of various sizes using the power-(log) normal model. In Recent Advances in Life Testing and Reliability (ed. Balakrishnan, N.), pp. 377–408, 1995.
- Rudin, W., Principles of Mathematical Analysis, McGraw-Hill, New York, 1976.
- Sarabia, J.M. and Castillo, E., About a class of max-stable families with applications to income distributions, Metron LXIII, 505–527, (2005).
- Tweedie, M.C.K., An index which distinguishes between some important exponential families, In Ghosh, J.K. and Roy, J. (ed.), Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, Indian Statistical Institute, Calcutta, pp. 579–604, 1984.
- Vaupel, J.W., Manton, K.G. and Stallard, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16, 439–54, (1979).